REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE** Journal of Educational Psychology	**3. DATES COVERED** *(From - To)* February 2010
4. TITLE AND SUBTITLE Posttraumatic Stress Disorder and Standardized Test-Taking Ability		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) L. Rutkowski, J.J. Vasterling, S.P. Proctor, C.J. Anderson		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute of Environmental Medicine, Kansas Street, Bldg. 42, Natick, MA 01760-5007		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702-5012		**10. SPONSOR/MONITOR'S ACRONYM(S)** M08-26
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Given the widespread use and high-stakes nature of educational standardized assessments, understanding factors that affect test-taking ability in young adults is vital. Although scholarly attention has often focused on demographic factors (e.g., gender and race), sufficiently prevalent acquired characteristics may also help explain widespread individual differences on standardized tests. In particular, this article focuses on the role that posttraumatic stress symptoms (PSS) potentially play in standardized academic assessments. Using a military sample measured before and after exposure to war-zone stressors, the authors sought to explain test-taking ability differences with respect to symptoms of PTSD on two cognitive tasks that closely match standardized test constructs. The primary method for this analysis is based on an item response theory with covariates approach. Findings suggest that the effect for PSS is significant on both tasks, particularly for those who experience the highest levels of PSS following war-zone exposure. Findings provide potentially valuable information regarding the nature of the relationship between PSS and verbal and logical reasoning test performance.

15. SUBJECT TERMS

Standardized testing, posttraumatic stress disorder, test taking

16. SECURITY CLASSIFICATION OF: unclassified			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Gabriele Furbay
a. REPORT unclassified	**b. ABSTRACT** unclassified	**c. THIS PAGE** unclassified	Unlimited	11	**19b. TELEPHONE NUMBER** *(include area code)* 508-233-4800

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Journal of Educational Psychology
2010, Vol. 102, No. 1, 223–233

0022-0663/10/$12.00 DOI: 10.1037/a0017287

Posttraumatic Stress Disorder and Standardized Test-Taking Ability

Leslie Rutkowski
VA Boston Healthcare System, Boston, Massachusetts; and University of Illinois at Urbana–Champaign

Jennifer J. Vasterling
VA Boston Healthcare System, Boston, Massachusetts; and Boston University School of Medicine

Susan P. Proctor
U.S. Army Institute of Environmental Medicine, Natick, Massachusetts; and VA Boston Healthcare System, Boston, Massachusetts

Carolyn J. Anderson
University of Illinois at Urbana–Champaign

Given the widespread use and high-stakes nature of educational standardized assessments, understanding factors that affect test-taking ability in young adults is vital. Although scholarly attention has often focused on demographic factors (e.g., gender and race), sufficiently prevalent acquired characteristics may also help explain widespread individual differences on standardized tests. In particular, this article focuses on the role that posttraumatic stress symptoms (PSS) potentially play in standardized academic assessments. Using a military sample measured before and after exposure to war-zone stressors, the authors sought to explain test-taking ability differences with respect to symptoms of PTSD on two cognitive tasks that closely match standardized test constructs. The primary method for this analysis is based on an item response theory with covariates approach. Findings suggest that the effect for PSS is significant on both tasks, particularly for those who experience the highest levels of PSS following war-zone exposure. Findings provide potentially valuable information regarding the nature of the relationship between PSS and verbal and logical reasoning test performance.

Keywords: standardized testing, posttraumatic stress disorder, test taking

As nearly every U.S. college applicant can attest, the majority of domestic colleges and universities require standardized tests for admission to undergraduate, graduate, and professional programs. Although controversial (Baron & Norman, 1992; FairTest, 2006), tests such as the Scholastic Aptitude Test (SAT), the Graduate Record Examination (GRE), and others are valued by higher educational institutions as predictors of first-year student grade point average (Bridgeman, McCamley-Jenkins, & Ervin, 2000) and graduate school success (Burton & Wang, 2005) and as an efficient measure of underlying traits such as math or reading ability. Given the widespread use and high-stakes nature of these assessments, understanding factors that affect test-taking ability in young adults is vital.

Factors that are largely determined by birth, such as gender and race, are important to any conversation about fair and equitable testing (for examples of these types of studies, see Arbuthnot, 2005; Holland, Hoffman, & Thompson, 2002; Ramist, Lewis, & McCamley-Jenkins, 1994; Schmitt & Dorans, 1990). Although scholarly attention has often focused on these birth factors, sufficiently prevalent acquired characteristics may also help explain widespread individual differences on standardized tests. In particular, this article focuses on the role that symptoms of posttraumatic stress disorder (PTSD) potentially play in academic assessments. PTSD is associated with symptoms such as intrusive thoughts, poor concentration, and hypervigilance to threat in the environment that could be predicted to interfere with test taking. More-

Leslie Rutkowski, VA Boston Healthcare System, Boston, Massachusetts, and Department of Educational Psychology, University of Illinois at Urbana–Champaign; Jennifer J. Vasterling, VA Boston Health Care System and Boston University School of Medicine; Susan P. Proctor, U.S. Army Institute of Environmental Medicine, Natick, Massachusetts, and VA Boston Health Care System; Carolyn J. Anderson, Department of Educational Psychology, University of Illinois at Urbana–Champaign.

The content of this article does not necessarily reflect the position or policy of the U.S. government, and no official endorsement should be inferred. This work was supported by awards from the U.S. Army Medical Research and Materiel Command (DAMD 17-03-0020; HSRRB Log No. A-11815) and VA Clinical Sciences Research and Development. The work was also supported in part by resources provided by the VA South Central Mental Illness Research, Education, and Clinical Center and the U.S. Army Research Institute for Environmental Medicine. The U.S. Army Medical Research Acquisition Activity (Fort Detrick, Maryland) is the awarding and administering acquisition office for DAMD 17-03-0020. Some of the work was completed at the Southeast Louisiana Veterans Healthcare System, New Orleans, Louisiana; the Department of Psychiatry and Neurology, Tulane University School of Medicine; the Department of Psychology, Tulane University; and the International Association for the Evaluation of Educational Achievement, Hamburg, Germany. We are especially grateful to the soldiers who donated their time to participate in the study and the efforts of the key military personnel who facilitated conduct of the study for their units. We also thank Dan and Lynda King for their valuable and constructive feedback.

Correspondence concerning this article should be directed to Leslie Rutkowski, who is now at the International Association for the Evaluation of Educational Achievement, Data Processing and Research Center, Mexikoring 37, 22297 Hamburg, Germany. E-mail: leslie.rutkowski@iea-dpc.de

over, a growing literature suggests that PTSD is associated with attention, working memory, and other cognitive deficits (Brewin, Kleiner, Vasterling, & Field, 2007; Hart et al., 2008; Vasterling & Brailey, 2005) that could likewise adversely affect performance on standardized academic tests.

A continuum of posttraumatic stress symptoms (PSS), including those sufficiently severe as to reach criteria for PTSD, may result from exposure to any extreme traumatic stressor such as military combat, physical and sexual assault, child abuse, disasters, or accidents (American Psychiatric Association [APA], 2000). In recent U.S. history, Hurricane Katrina, Operation Iraqi Freedom (OIF)/Operation Enduring Freedom, and the terrorist attacks of September 11, 2001, typify events that might trigger PSS or PTSD. Current diagnostic classifications group PTSD symptoms into three clusters: (a) reexperiencing of the traumatic event (e.g., nightmares, intrusive thoughts), (b) avoidance of stimuli associated with the traumatic event and numbing of general responsiveness (e.g., restricted range of affect, loss of interest in previously engaging activities), and (c) increased arousal symptoms (e.g., poor concentration, sleep disturbance).

Many Americans are exposed to inner-city violence, family violence, rape, and other extreme stress. The National Comorbidity Study–Replicate (Kessler et al., 2005) estimated the lifetime prevalence of PTSD in a nationally representative community-based sample to be 6.8%. The prevalence of PSS and PTSD may be even higher in at-risk populations such as war-zone veterans. For example, according to a major study of Vietnam-era veterans (Kulka et al., 1990), nearly a third of men (30.9%) and over a quarter of women (26.9%) who served in Vietnam experienced PTSD at some point in their lives, with an additional 22.5% of men and 21.5% of women experiencing a subset of PTSD symptoms that were notable but not sufficient to meet full diagnostic criteria. Reanalysis of a male-only subset of the same Vietnam sample but with more stringent diagnostic criteria found an adjusted lifetime PTSD figure of 18.9% (Dohrenwend et al., 2006). Nonpopulation-based samples of OIF veterans have revealed screening-based estimates of PTSD that range from 11.6% to 12.9% among recently returned military personnel (Hoge et al., 2004; Vasterling et al., 2006), with rates increasing over time (Milliken, Auchterlonie, & Hoge, 2007). A combined sample of U.S service members deployed to Iraq or Afghanistan demonstrated screening-based PTSD rates of 13.8% (Schell & Marshall, 2008), with new onset rates of 7.6% among combat-exposed study participants (Smith et al., 2008). The prevalence of PTSD among the groups outlined above suggests that if deleterious effects of PTSD on test-taking ability are found, a large group of people could be at a significant disadvantage in testing situations used for promotion or college admission.

Several studies have examined cross-sectional relationships between chronic PTSD and performance on the types of constructs measured in standardized assessments for college admissions, finding that IQ scores are inversely related to PTSD symptom severity (Brandes et al., 2002; Gil, Calev, Greenberg, Kugelmass, & Lerer, 1990; Gilbertson, Gurvits, Lasko, Orr, & Pitman, 2001; Gurvits et al., 2000, 1993; Vasterling, Brailey, Constans, Borges, & Sutker, 1997; Vasterling et al., 2002). In particular, Brandes et al. (2002) and Vasterling et al. (2002) found Pearson correlations of approximately $-.30$ between measures of PTSD symptoms and measures of intelligence. In a study relying in part on archival military data, Macklin et al. (1998) likewise found that current intellectual performance was inversely related to PTSD symptom severity with a partial correlation of $-.37$. However, cross-sectional associations between postcombat measures of current intelligence and PTSD symptom severity were no longer significant after controlling for precombat intelligence estimated from archival records, suggesting that precombat intelligence may have created additional risk of PTSD, rather than PTSD affecting intellectual performance.

The literature examining the relationships between exposure to violence (a common predictor of PTSD) and the academic achievement of adolescents and adults also suggests an association between traumatic experiences and standardized test performance. For example, Schwab-Stone et al. (1995) documented a significant negative relationship between direct exposure to violence and school achievement in a sample of over 2,000 adolescents in an urban community. A similar study (Schwartz & Gorman, 2003) found a negative relationship between exposure to community violence and academic functioning as measured by a standardized test of achievement and grade point average. Examining associations specifically between PTSD and achievement in Lebanese adolescents exposed to frequent occurrences of violence such as terrorist attacks and artillery fire, Saigh, Mroueh, and Bremner (1997) found that those adolescents with PTSD, compared with those without PTSD, had lower levels of scholastic achievement on the Metropolitan Achievement Test, a standardized index of academic achievement in the areas of reading, mathematics, and language.

Missing from the literature, however, is prospective research allowing greater inferences regarding the potential causal pathway between PSS and test taking. The current study uses prospectively gathered data from the Neurocognition Deployment Health Study (NDHS; Vasterling et al., 2006) to examine potential changes in test-taking ability as a function of PSS. The study from which the data are drawn included neurocognitive and emotional assessment of a cohort of 1,595 U.S. Army soldiers, many of whom eventually deployed to Iraq in the support of OIF. Some of the neurocognitive tasks administered in this study evaluated processes similar to those measured in a standardized testing environment. Relevant measures include tasks assessing logical reasoning and vocabulary—cognitive skills measured on standardized tests such as the SAT, ACT, and GRE. The availability of both pre- and postwar-zone neurocognitive and PTSD symptom data makes the data set uniquely suited to examine the effects of PSS on standardized test performance.

Because the current study targets how the examinee's ability to correctly answer a standardized test question is affected by the acquired characteristic of PSS, item response theory (IRT) models were fit to the data. In IRT, responses to items are viewed as observable indicators of an individual's latent ability in which all examinees (and items) can be placed on a common scale to assess how much of the latent trait an examinee has and how much of the latent trait an examinee needs to correctly answer items with high probability. The current analysis uses IRT with covariates (Adams, Wilson, & Wu, 1997; de Boeck & Wilson, 2004; Zwinderman, 1991, 1997) to examine the relationship between PSS and test-taking ability as evidenced by responses to two tasks that tap skills similar to those measured on standardized tests. To control for the possibility that study participants suffered from PSS prior to OIF

war-zone exposure, we included an effect for predeployment PSS in the model.

The addition of covariates in the traditional IRT model is particularly useful for flexible modeling of categorical survey and assessment data and for explaining individual differences. Although IRT is typically limited to descriptive uses (as outlined above), IRT with covariates is also useful for explanatory purposes. Given this latter use, the study can investigate a possible relationship between responses to items on a test and other variables related to the item or the examinee (for a detailed discussion of IRT, see de Boeck & Wilson, 2004; Embretson & Reise, 2000).

Based on the potential for PSS and associated cognitive impairment to interfere with standardized tests, it was hypothesized that standardized test performance would be negatively affected by the acquisition or exacerbation of PSS. Specifically, we predicted that after taking into account baseline standardized test scores, combat experience, and baseline levels of PSS, postdeployment PSS would be negatively associated with vocabulary and reasoning test-taking ability. Findings provide potentially valuable information regarding the nature of the relationship between PSS and verbal and logical reasoning test performance.

Method

Study Design and Sampling

Participants were drawn from the larger NDHS study sample. The current study included only those from the larger cohort (N = 654) who (a) were active-duty Army soldiers, (b) deployed to the Iraq war zone during the first wave of NDHS data collection, and (c) completed predeployment assessments (Time 1, between April and December 2003) and postdeployment assessments (Time 2, between January and May 2005). In the larger study, sampling was conducted at the battalion level, with battalions chosen to reflect heterogeneous deployment experiences (Vasterling et al., 2006). Based on power calculations and anticipated participation and attrition rates, a target sample size of 850 deploying soldiers was selected for the larger study. Participants, referred at random to the study by battalion commanders, consented individually and were offered a way to exit the study area unobserved if they declined to participate. At the individual level, exclusion criteria included pending separation from military service or reassignment or physical limitations.

Sample Characteristics

Sample demographic characteristics can be found in Table 1. By occupational specialty our sample was as follows: infantry (n = 234), maintenance (electronics and mechanical; n = 152), communications and intelligence (n = 101), health care (n = 43), support and administration (n = 43), supply (n = 54), other (n = 27). In assessing the occupational distribution of the sample, it is important to consider that OIF has been characterized by high levels of combat exposure, even in traditionally noncombat occupational specializations. The proportion of participants experiencing several types of combat experiences, as measured by a modified version of the Combat Experiences module of the Deployment Risk and Resilience Inventory (King, King, & Vogt, 2003), are included in Table 2.

Table 1
Descriptive Statistics for Sample Used in Current Study

Descriptive	N	%	Minimum	Maximum	M	SD
Age	654		17.68	46.48	25.03	5.25
Gender	653					
Female	56	8.56				
Male	597	91.28				
Highest grade level (school)	653		8.00	18.00	12.46	1.25
Years in the Army	653		0.00	24.00	3.91	4.26
Marital status	654					
Single	305	46.64				
Married	297	45.41				
Divorced/separated	47	7.19				
Live-in partner	5	0.76				
Gender (% male)	600	91.70				
Race/ethnicity	654					
African American	106	16.21				
Asian American	17	2.60				
Caucasian	369	56.42				
Hispanic American	96	14.68				
Other	66	10.09				
Assessment scores						
Time 1 logical reasoning	654		1	24	20.83	3.61
Time 2 logical reasoning	654		3	24	21.30	3.33
Time 1 vocabulary	654		3	25	16.10	5.09
Time 2 vocabulary	654		3	25	16.85	5.04
PCL-C score						
Time 1	654		17	78	29.16	12.51
Time 2	654		17	80	32.33	13.21

Note. PCL-C = Posttraumatic Stress Disorder Checklist, Civilian version.

Table 2
Percentage of Study Participants With Combat Experience During Deployment

Combat experience	*N*	Ever	*SE*	At least a few times per week	*SE*
Went on combat patrols or missions	651	91	1.1	61	1.9
Encountered land or water mines and/or booby traps	647	60	1.9	19	1.5
Received hostile incoming fire from small arms, artillery, rockets, mortars, or bombs	649	98	0.6	67	1.8
Received friendly incoming fire from small arms, artillery, rockets, mortars, or bombs	649	22	1.6	4	0.8
In a vehicle that was under fire	652	74	1.7	23	1.7
Attacked by terrorists or civilians	646	69	1.8	26	1.7
Part of a land or naval artillery unit that fired on the enemy	648	23	1.6	8	1.1
Part of an assault on entrenched or fortified positions	648	32	1.8	6	0.9
Took part in an invasion that involved naval and/or land forces	645	29	1.8	6	0.9
In a unit that engaged in battle in which it suffered casualties	648	64	1.9	8	1.1
Witnessed someone from own unit or an ally unit being seriously wounded or killed	649	55	2.0	4	0.7
Witnessed soldiers from enemy troops being seriously wounded or killed	650	61	1.9	9	1.1
Was wounded or injured in combat	650	14	1.4	0	0.3
Fired weapon at the enemy	651	60	1.9	15	1.4
Killed or thought killed someone in combat	649	44	2.0	5	0.9
Participated in a support convoy	650	95	0.9	37	1.9

It is notable that the majority (61%) of participants were involved in combat patrols or missions at least a few times per week. Further, of those who were involved in combat patrols or missions at least a few times per week, 64%, $\chi^2(1) = 6.64$, $p = .01$, indicated that they also received hostile incoming fire from small arms, artillery, rockets, mortars, or bombs at least a few times per week. Regarding past deployment history, of those sampled for the current article, 14 had deployed at least once to a hazardous area[1] excluding the current deployment since 2001. Only two participants had deployed twice to a hazardous area since 2001.

Measures

Posttraumatic Stress Disorder Checklist, Civilian version (PCL-C). The PCL-C (Weathers, Huska, & Keane, 1991) is a widely used, 17-item self-report scale that measures the severity of PTSD symptoms. Respondents are asked to indicate how much PTSD symptoms have bothered the respondent on a 5-point Likert scale (from *not at all* to *extremely*), without reference to a specific traumatic experience. Possible scores on the PCL-C range from 17 (all responses are *not at all*) to 85 (all responses are *extremely*). Items on the PCL-C are congruent with the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; APA, 2000) and address each of the three symptom clusters. For example, respondents are asked how often "they feel distant or cut off from people" or how often they "have repeated, disturbing dreams of a stressful military experience." Time 1 and Time 2 Cronbach's alphas for the PCL-C are .93 and .94, respectively. Other studies have found the PCL to be characterized by high test–retest reliability (*rs* = .92 and .88, immediate and 1-week retest, respectively), internal consistency ($\alpha = .94$), and convergent validity (*rs* > .75) with other PTSD measures (Ruggiero, Del Ben, Scotti, & Rabalais, 2003). Further, the PCL was found to correlate well with the Clinician-Administered PTSD Scale ($r = .93$), and it is recommended as a good screening and self-report measure of PTSD (Blanchard, Jones-Alexander, Buckley, & Forneris, 1996). In the current sample, women scored about four points higher on Time 1 PCL than men, $t(649) = 2.55$, $p < .01$. No gender differences existed at Time 2. Older participants had, on average, lower PCL scores at Time 1. For each 5-year increase in age, participants scored about one and a half points lower on the PCL, $t(651) = 3.21$, $p < .01$. No age differences existed at Time 2.

Automated Neuropsychological Assessment Metrics (ANAM) logical reasoning assessment. The ANAM logical reasoning task (Reeves, Kane, Elsmore, Winter, & Bleiberg, 2002) measures grammatical and logical reasoning. The logical reasoning task is taken from the larger ANAM battery, a clinical battery originally designed by the Office of Military Performance Assessment Technology to measure cognitive functioning across administrations (Kabat, Kane, Jefferson, & DiPino, 2001). The larger assessment has proven useful in a number of clinical applications and as a cost-effective measure of cognitive function (Jones, Loe, Krach, Rager, & Jones, 2008). The accuracy measure of the logical reasoning task has been found to correlate exceptionally well with the Cognitive Efficiency cluster of the Woodcock–Johnson Tests of Cognitive Ability (Jones et al., 2008). All 24 logical reasoning items present both a logical rule (such as *& comes before #*) and a logical relation (such as *& #*) in which the examinee chooses whether the relation is the same as or different from the rule. (In the previous example the correct answer is *same*; that is, & does come before #, as stated in the rule.) In this sample, the reliabilities were $\alpha = .84$ for Time 1 and $\alpha = .82$ for Time 2. No age or gender differences existed on this measure at either time point.

NES3 vocabulary assessment. The NES3 vocabulary task is a computer-administered 25-item multiple-choice test designed to estimate general verbal ability (Letz, 2000) and is derived in part from the Armed Forces Qualification Test–Verbal subtest. The larger NES3 assessment is designed to assess neurobehavioral function in studies of environment and occupational health. The NES3 vocabulary task correlates well with the Wechsler Adult Intelligence Scale–Revised vocabulary test (Krengel et al., 1996). In this sample, the reliabilities were $\alpha = .87$ for both Times 1 and 2. No gender differences existed on this measure at Time 1 or Time 2. Older participants, on average, scored higher on the vocabulary

[1] For the current article, *hazardous area* is defined as Afghanistan, Iraq, Bosnia, Kosovo, or Kuwait.

task. A 5-year age increase equated to approximately a one-point increase in a participants' vocabulary score at Time 1, $t(648) = 5.91$, $p < .01$, and approximately a one-point increase at Time 2, $t(651) = 5.22$, $p < .01$.

Combat experiences. Combat exposure was measured with the Combat Experiences scale from a modified version of the Deployment Risk and Resilience Inventory (King et al., 2003). The Combat Experiences scale is a 15-item, five-category Likert scale. Response options range from the experience *never* happened to the experience happened *daily or almost daily*. Higher sum scores on this scale are indicative of greater combat exposure. A complete list of these items can be found in Table 2. In this sample, internal consistency was high ($\alpha = .90$).

Other covariates. To control for other person and contextual factors that may also be responsible for changes in test-taking ability, we added a number of covariates to the model. These include age, gender, average number of hours of sleep for the week prior to Time 2 data collection assessment and average weekly alcohol consumption (in number of drinks) for the month prior to the Time 2 data collection. Finally, given the association between PTSD and traumatic brain injury ($OR = 2.98$, 95% CI [1.70, 5.24]; Hoge et al., 2008), we included as a predictor whether or not the respondent reported a head injury resulting in loss of consciousness between the pre- and postdeployment data collections.

Analysis Method

In the current study, we used a latent regression Rasch model (Adams et al., 1997; de Boeck & Wilson, 2004; Zwinderman, 1991, 1997) that included attributes of the person to explain individual differences. This method permitted the addition of covariates in IRT models. The latent regression Rasch model is a type of multilevel IRT model that has been shown to have utility in analyzing item response data when explaining individual differences is of interest (Cheong & Raudenbush, 2000; Pastor, 2003). The power of the latent regression Rasch model is in the addition of predictors that allow for a flexible exploration of individual differences with respect to (latent) ability, which standardized tests are presumed to measure. Specifically, adding covariates for PSS and a number of control variables into the Rasch model allowed an examination of possible associations between PSS and an examinee's test-taking ability.

The latent regression Rasch model is an extension of a standard Rasch model (Rasch, 1980) with the addition of linear predictors for the person's value on the latent trait. The model for the latent trait, θ_p, is a linear regression equation; that is,

$$\theta_p = \sum_{j=1}^{J} \gamma_j Z_{pj} + \varepsilon_p, \tag{1}$$

where Z_{pj} is the value of covariate j ($j = 1, \ldots, J$) for person p and γ_j is the regression coefficient of covariate j. This model includes a random-person effect, ε_p, which represents unexplained variability between people in terms of their ability. In line with typical IRT conventions, the latent trait is standardized to a mean of zero and a standard deviation of one. The item difficulty parameters in the latent regression Rasch model β_i are identical to the item difficulty parameters in the classical Rasch model. Using this approach yields the following model for responses to items:

$$P(Y_{ip} = 1 | \theta_p, \beta_i) = \frac{\exp\left(\sum_{j=1}^{J} \gamma_j Z_{pj} + \varepsilon_p - \beta_i\right)}{1 + \exp\left(\sum_{j=1}^{J} \gamma_j Z_{pj} + \varepsilon_p - \beta_i\right)}, \tag{2}$$

where $P(Y_{ip} = 1 | \theta_p, \beta_i)$ is the probability that a person with ability θ_p gives a correct response on item i with difficulty β_i.

All IRT models were fit to data with PROC NLMIXED in SAS 9.1 (SAS Institute, 2003; for examples of input code, see de Boeck & Wilson, 2004; Sheu, Chen, Su, & Wang, 2005). Likelihood ratio tests were used to test the significance of the effect of PSS (for a discussion of likelihood ratio tests, see Agresti, 2007).

Models

To investigate whether changes in PSS during deployment significantly predicted differences in examinees' ability to correctly answer test items, we fit several IRT models to the data. All models were fit twice: once each for the Time 2 logical reasoning and the vocabulary item responses. To create a measure of residualized change taking into account Time 1 values of PSS and cognitive task scores, in every model, we included Time 1 PSS and the Time 1 (predeployment) value of the relevant cognitive task score as predictors. In addition, we included a number of covariates to control for other factors that may contribute to ability differences. The covariates, taken from Time 2 measurements, included combat exposure, gender, age, alcohol consumption, sleep, and head injury with loss of consciousness. Time 2 PSS measures entered into the model either as the total PCL score or by symptom cluster. Given the high level of collinearity between symptom clusters, subscale scores were not entered into a single model. Rather, for those models that examined the effect of symptom cluster on test-taking ability, each symptom cluster score was used separately as a Time 2 predictor. This approach resulted in four models each for the logical reasoning and the vocabulary items. The effects estimated in the vocabulary and logical reasoning models are detailed in Table 3.

Table 3
Summary of Predictors for Each Model

Effect	Model 1	Model 2	Model 3	Model 4
PSS[a]	x	x	x	x
PSS[b]	x			
Reexperiencing[b]		x		
Avoidance/numbing[b]			x	
Hyperarousal[b]				x
Head Injury[b]	x	x	x	x
Combat experience[b]	x	x	x	x
Age[b]	x	x	x	x
Gender[b]	x	x	x	x
Sleep[b]	x	x	x	x
Alcohol[b]	x	x	x	x
Cognitive assessment score[a]	x	x	x	x

Note. Effects are identical for the logical reasoning and vocabulary models. PSS = posttraumatic stress symptoms.
[a] Measured at Time 1. [b] Measured at Time 2.

The model fit to the NDHS data was

$$P(Y_{ip} = 1|\theta_p, \beta_i^*) = \frac{\exp\left(\gamma_1 PSS1_p + \gamma_2 PSS2_p + \sum_{j=3}^{J} \gamma_j Z_{pj} + \varepsilon_p - \beta_i\right)}{1 + \exp\left(\gamma_1 PSS1_p + \gamma_2 PSS2_p + \sum_{j=3}^{J} \gamma_j Z_{pj} + \varepsilon_p - \beta_i\right)}, \quad (3)$$

where γ_1 was the coefficient for the effect of PSS before deployment, γ_2 was the coefficient for the effect of PSS or symptom cluster after deployment, γ_j were the other covariates as listed in the Measures section, and β_i was the item difficulty.

Significant parameter estimates for Time 2 PSS (i.e., γ_2) suggest that as an individual's PCL score or symptom cluster score changes, the probability of correctly answering an item changes according to the level of symptom severity. In other words, a significant negative effect for Time 2 PSS suggests that this disorder reduces test-taking ability. Given the prospective design of the study, we can reasonably attribute this reduction to changes in PSS.

Results

The parameters from the models fit to the logical reasoning data and the vocabulary data are presented in Table 4.

The logical reasoning models all yielded significant likelihood ratio test statistics for PSS symptoms at Time 2: PSS, $\chi^2(1) = 84$, $p < .01$; reexperiencing, $\chi^2(1) = 19$, $p < .01$; avoidance–numbing, $\chi^2(1) = 18$, $p < .01$; hyperarousal, $\chi^2(1) = 85$, $p < .01$. That is, for each model, the addition of PCL scores or subscale scores at Time 2 significantly increased the fit of the relevant model when compared with a model that did not contain the effect for Time 2 PCL scores. For the logical reasoning model in which Time 2 PCL total score was entered into the model, the effect of PSS at Time 2 was significant when controlling for PSS at Time 1 and the other covariates, $\hat{\gamma}_2 = -.01$, $t(636) = -5.15$, $p < .01$. Further, with the exception of Time 1 logical reasoning performance, Time 2 PSS symptoms was the only significant effect in the model. In other words, none of the other covariates that might be associated with diminished ability were significantly associated with correctly answering logical reasoning items.

On average, study participants reported a Time 2 PCL score of 32.33 (SD = 13.21). This suggests that when holding all other covariates constant, the probability of correctly answering the average logical reasoning item for someone with a PCL score of 32 is approximately .61. In comparison, the probability of a correct answer for a person with a PCL score of 17 (the lowest possible score) is .64. This suggests an approximately 3% average reduction in the probability of correctly answering the average logical reasoning item at Time 2 for a study participant with an average PCL score. Although the average effect was small, at the extreme end of the range, the effect was much larger. For example, a person

Table 4
Summary of Models

	Model 1				Model 2				Model 3				Model 4			
Model parameter	Estimate	*SE*	*df*	*t*	Estimate	*SE*	*df*	*t*	Estimate	*SE*	*df*	*t*	Estimate	*SE*	*df*	*t*
Logical reasoning																
PSS[a]	.00	.00	636	0.23	.00	.00	635	−0.12	.00	.00	635	−0.18	.00	.00	634	0.01
PSS[b]	−.01*	.01	636	−5.15												
Reexperiencing[b]					−.01	.01	635	−1.63								
Avoidance/numbing[b]									−.01	.01	635	−1.41				
Hyperarousal[b]													−.01	.01	634	−1.72
Head injury[b]	.00	.01	636	0.17	.00	.01	635	0.09	.00	.01	635	0.05	.00	.01	634	0.13
Combat experience[b]	.02	.11	636	−0.22	.02	.11	635	0.16	.00	.11	635	0.04	−.03	.11	634	−0.26
Age[b]	.03	.02	636	1.45	.03	.02	635	1.44	.03	.02	635	1.46	.03	.02	634	1.38
Gender[b]	.00	.00	636	1.48	.00	.00	635	1.46	.00	.00	635	1.41	.00	.00	634	1.48
Sleep[b]	.05	.11	636	0.49	.05	.11	635	0.46	.04	.11	635	0.40	.06	.11	634	0.49
Alcohol[b]	.00	.00	636	−1.63	.00	.00	635	−1.62	−.01	.00	635	−1.80	−.01	.00	634	−1.84
Logical reasoning[a]	.05*	.00	636	27.96	.05*	.00	635	27.49	.05*	.00	635	27.38	.05*	.00	634	27.55
Vocabulary																
PSS[a]	.00	.00	637	0.30	.00	.00	636	−0.39	.00	.00	636	−0.02	.00	.00	635	−0.10
PSS[b]	−.01*	.00	637	−5.50												
Reexperiencing[b]					−.01	.00	636	−1.75								
Avoidance/numbing[b]									−.01	.00	636	−2.50				
Hyperarousal[b]													−.01	.00	635	−1.97
Head injury[b]	.01	.00	637	1.78	.01	.00	636	1.68	.01	.00	636	1.73	.01	.00	635	1.67
Combat experience[b]	.03	.07	637	0.38	.01	.07	636	0.08	.00	.07	636	−0.01	.02	.07	635	0.26
Age[b]	.00	.01	637	0.17	.01	.01	636	0.49	.01	.01	636	0.41	.00	.01	635	0.23
Gender[b]	.00	.00	637	−0.25	.00	.00	636	−0.35	.00	.00	636	−0.33	.00	.00	635	−0.32
Sleep[b]	.01	.07	637	0.10	.00	.07	636	0.07	.00	.07	636	0.00	.01	.07	635	0.08
Alcohol[b]	.00	.00	637	−1.45	.00	.00	636	−1.76	.00	.00	636	−1.72	.00	.00	635	−1.69
Vocabulary[a]	.22*	.00	637	46.51	.22*	.00	636	46.29	.22*	.00	636	46.33	.22*	.00	635	46.44

Note. PSS = posttraumatic stress symptoms.
[a] Measured at Time 1. [b] Measured at Time 2.
* $p < .05$ (Bonferroni adjusted < 0.013).

with a Time 2 PCL score of 71 (the maximum score among NDHS participants) would have an 11% lower probability of correctly answering the most difficult logical reasoning than someone with the lowest PCL score. Even more pronounced were the differences in correct response probabilities on a logical reasoning item of average difficulty. For an item of this type, the difference in the probability of a correct answer was more than 13% between those with the lowest Time 2 PSS levels (PCL = 17) and the highest Time 2 PSS levels (PCL = 71).

Figure 1 displays item characteristic curves for a logical reasoning item of average difficulty and participant groups with the lowest versus the highest Time 2 PCL scores. The gray curve represents correct response probabilities for participants with the lowest observed Time 2 PCL scores, and the black curve denotes correct response probabilities for participants with the highest observed Time 2 PCL scores. Here we can see that regardless of ability level, the probability of correctly answering a typical logical reasoning item is lower for the group with the highest level of PSS at Time 2. Only seven participants in the sample were at this pathological extreme, whereas 28 participants reported a Time 2 PCL score of 60 points or more.

For the logical reasoning models in which PCL symptom cluster scores were entered into the model, the effects of the symptom cluster scores measured at Time 2 were not significant:[2] reexperiencing, $\hat{\gamma}_2 = -.01$, $t(635) = -1.63$, $p = .10$; avoidance–numbing, $\hat{\gamma}_2 = -.01$, $t(635) = -1.41$, $p = .16$; hyperarousal, $\hat{\gamma}_2 = -.01$, $t(634) = -1.72$, $p = .09$. Indeed, besides the effect of Time 1 cognitive assessment scores, there were no significant effects in the subscale models for the logical reasoning items. This suggests that no single Time 2 PSS cluster was responsible for differences in logical reasoning test-taking ability. Rather, findings suggest that the full spectrum of PTSD symptoms was responsible for logical reasoning ability differences.

As with the logical reasoning models, the vocabulary models also yielded significant likelihood ratio test statistics for PSS symptoms at Time 2: PSS, $\chi^2(1) = 93$, $p < .01$; reexperiencing, $\chi^2(1) = 52$, $p < .01$; avoidance–numbing, $\chi^2(1) = 55$, $p < .01$; hyperarousal, $\chi^2(1) = 130$, $p < .01$. The results suggest that in each vocabulary model, the fit was significantly improved by adding an effect for Time 2 PSS. In terms of significant effects in the models fit to the data, the findings were similar to the logical reasoning models. That is, the vocabulary model that included Time 2 PCL scores exhibited a significant Time 2 PSS effect when controlling for the other effects in the model, $\hat{\gamma}_2 = -.01$, $t(647) = -5.00$, $p < .01$. Besides the Time 1 vocabulary assessment score, none of the other predictors in the model were significant.

The Time 2 PSS effect can be interpreted such that given an average respondent with a Time 2 PCL score of 32 (the average PCL score in the sample), we expect that the probability of a correct answer on an average vocabulary item is .55. Compared with that of a respondent whose Time 2 PCL score is just 17 and an associated probability of a correct answer at about .58, there is a 3% higher probability of an incorrect answer from the respondent with a higher Time 2 PCL score, all else equal.

Vocabulary test-taking ability differences were more pronounced when comparing study participants at the highest and lowest end of the Time 2 PCL spectrum on the hardest items. For example, a respondent with the highest Time 2 PCL score would correctly answer the most difficult vocabulary item about 6% of the time. In comparison, a respondent with the lowest Time 2 PCL score would answer the same item correctly about 11% of the time, for a difference of 5%. However, the largest disparity between study participants at the high and low end of the Time 2 PSS spectrum was on items of average difficulty, where differences in the probability of correct answers emerged on the order of more than .13, $P(Y_p = 1|\theta_p^{HIGH\ PSS}, \bar{\beta}) = .45$ versus $P(Y_{ip} = 1|\theta_p^{LOW\ PSS}, \bar{\beta}) = .58$. The differences in probabilities of a correct response to an average vocabulary item are presented in Figure 2. Although on average the difficulty for vocabulary items was higher than for logical reasoning items, we found similar results on the vocabulary assessment between those with the highest PCL scores and those with the lowest PCL scores. That is, across the ability continuum, those with the highest levels of PSS had lower probabilities of a correct response.

Similar to the logical reasoning models, the vocabulary models in which Time 2 PCL symptom cluster scores were entered into the model did not show significant Time 2 PSS effects with an adjusted significance level: reexperiencing, $\hat{\gamma}_2 = -.01$, $t(636) = -1.75$, $p = .08$; avoidance–numbing, $\hat{\gamma}_2 = -.01$, $t(636) = -2.50$, $p = .013$; hyperarousal, $\hat{\gamma}_2 = -.01$, $t(635) = -1.97$, $p = .05$. Again, there were no significant effects in the subscale models for the vocabulary items other than the Time 1 cognitive assessment scale. These findings substantiate the results from the logical reasoning models. That is, individual symptom clusters were not sufficient to diminish test-taking ability. Instead, Time 2 PSS, as measured by all three symptom clusters, seems to be an important determinant of vocabulary test-taking ability following exposure to an extreme traumatic stressor.

Discussion

In this article, we used IRT with covariates to assess whether changes in PTSD symptomatology had a significant effect on test-taking ability on two cognitive tasks administered after exposure to wartime stressors that measure constructs similar to those assessed on standardized tests. Findings indicated that for both the logical reasoning task and the vocabulary task, a residualized measure of Time 2 PSS adjusted for Time 1 PSS values was significantly associated with diminished ability to answer items correctly, especially for participants who showed the largest increase in PSS at Time 2. At the extreme, people with the highest levels of Time 2 PSS would face a 13% reduction in the probability of correctly answering a typical logical reasoning item or vocabulary item when compared with those with the lowest Time 2 PSS levels.

Previous research on college-age groups suggests that educational attainment is negatively impacted by anxiety disorders (Kessler, Foster, Saunders, & Stang, 1995); however, less is known about the specific effects of anxiety disorders on test-taking ability, particularly from a prospective approach. The current study sheds light on this issue and suggests that after controlling for predeployment PSS and a number of possibly confounding factors, PTSD symptoms adversely affect test-taking ability in study participants, and that there is a dosing effect in which more severe symptoms are associated with poorer test taking.

[2] Given that four models were fit to the data, we used a Bonferroni adjusted significance level, $\alpha/C = .05/4 = .013$, where C is equal to the number of hypotheses tested.

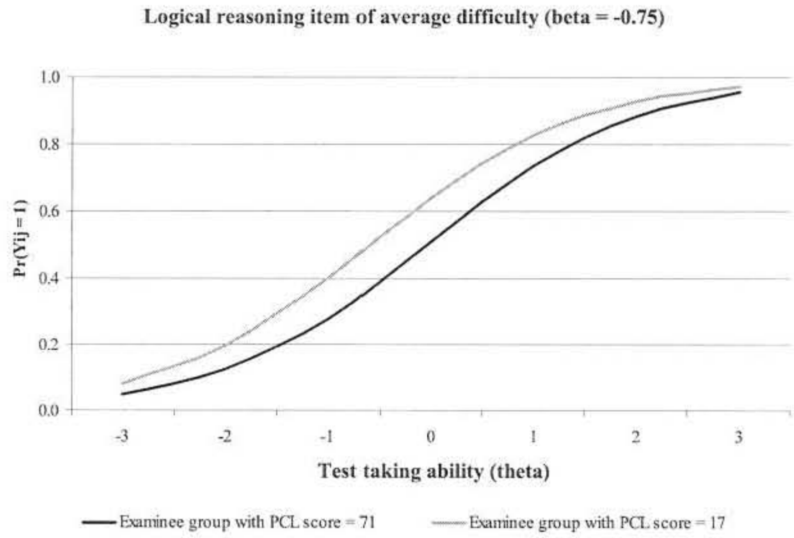

Figure 1. Item characteristic curves for a logical reasoning item of average difficulty and two groups with the highest and lowest levels of posttraumatic stress symptoms. PCL = Posttraumatic Stress Disorder Checklist.

To interpret the possible effect of the highest level of PSS on total test scores, we simulated item response data for 2,000 examinees of average ability, 1,000 each in the low- and high-PSS groups, corresponding to PCL scores of 17 and 71, respectively. Using item difficulties calculated in earlier analyses, latent trait values of zero for those with the lowest levels of PSS and adjusted latent trait values of $-.71$ for those with the highest levels of PSS, we generated item responses using our IRT model for all 2,000 examinees on both cognitive tasks. This resulted in item responses for 2,000 examinees on 24 items for the logical reasoning assessment and 25 items for the vocabulary assessment. On the basis of this method, we found that those at the lowest end of the PSS spectrum received an average score of 16.07 on the logical reasoning assessment, whereas those in the high-PSS group received a significantly lower average score of 11.80, $M_{\text{diff}} = 4.27$, $t(998) = -4.20$, $p < .01$. Findings were similar for the vocabulary assessment, for which simulated data resulted in a low-PSS group mean of 17.74 and, again, a significantly lower mean for the high-PSS group of 12.69, $M_{\text{diff}} = 5.05$, $t(998) = -5.06$, $p < .01$. Score differences on both cognitive assessments between those in the low- and high-PSS groups also suggest meaningful practical differences as indicated by large Cohen's d effect sizes (logical reasoning, $d = 1.66$; vocabulary, $d = 1.90$). These findings indicate that widespread test-taking ability differences stemming from PSS can have important consequences on cognitive assessment scores.

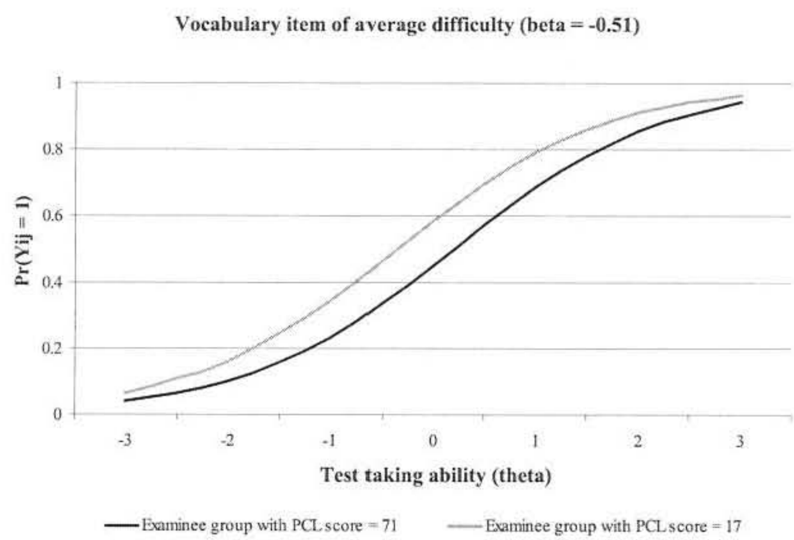

Figure 2. Item characteristic curves for a vocabulary item of average difficulty and two groups with the highest and lowest levels of posttraumatic stress symptoms. PCL = Posttraumatic Stress Disorder Checklist.

Given the significant effect that Time 2 PSS has on an examinee's ability to correctly answer the two cognitive tasks used in this study, it is reasonable to expect that these findings may be relevant in other contexts. As of the end of 2007, more than 1.64 million service members have deployed in support of the wars in Iraq and Afghanistan, with some units serving multiple rotations of 12 to 15 months (Tanielian & Jaycox, 2008). It is important to consider that many of these military servicewomen and -men will pursue higher education or otherwise face testing situations for promotion or job placement. Estimates suggest that Montgomery GI Bill usage rates exceed 65% (Winter, 2005). Understanding how veterans' experiences impact on their ability to pursue higher education or career advancement is important for both the mental health and the education communities.

If indeed this article's findings do generalize to a civilian population, the implications for this research may be far-reaching. Estimates suggest that over 125,000 children in New Orleans were displaced as a result of Hurricane Katrina (Redlener, 2006), and nearly one half of children in shelters exhibit some type of emotional or behavioral disorder such as PTSD (Abramson & Garfield, 2006). Internationally, PTSD rates among children are estimated at 10% in Baghdad (Eccleston, 2007) and nearly 33% in Mosul, Iraq (Eccleston, 2007), and 13% in posttsunami southern Thailand (Thienkrua et al., 2006). Internationally, areas such as these are recipients of recovery aid from international organizations that commonly mandate adherence to structural adjustment programs, a component of which may include standardized tests of achievement as markers of sufficient progress. Our findings suggest that in this context, achievement results from standardized cognitive assessments should be used with caution, if at all. Alternatively, test administrations in known conflict or disaster areas should include a PTSD scale so that proficiency score estimates can be adjusted accordingly. Either empirical estimates of the PTSD effect can be used or additional studies regarding the magnitude of the PTSD effect on ability could be undertaken.

There are several limitations associated with the study. First, although the median age (23.5 years) of the cohort is fairly representative of the median age (20.5 years) of U.S. college students (National Center for Education Statistics, 2005), the proportion of women (8%) is not representative. Furthermore, systematic rather than population-based sampling was used to derive the study sample and included only one service branch, which limits the generalizability of the findings to a broader population. Although 27% of the study participants reported education levels beyond high school, participants may differ systematically from young adults who choose college over military service. Furthermore, cognitive assessments used for the current analysis were drawn from those collected during the NDHS, and they do not represent the exact types of items found on standardized college entrance tests such as the SAT and the GRE. However, the cognitive processes examined closely match many of those measured on standardized assessments. Also, we did not assess clinical PTSD diagnoses.

Regarding the tasks used in this analysis, the items may have been too easy to fully detect differences associated with PSS. Item difficulties were in general quite low and ranged from approximately -3.91 to 1.97 at Time 2. Similarly, the probability that an average examinee would correctly answer an average item ranged from a low of approximately .71 to nearly .96, depending on the cognitive assessment. As a result, findings may underestimate the impact of PSS on test-taking ability; however, the characteristics of the logical reasoning and vocabulary tasks do allow some additional insights into the effect of PSS on test-taking ability. That is, the relative ease of the task and low-stakes nature of the testing context suggest that processes other than simple test anxiety explain associations between test performance and PSS (Hembree, 1988).

The findings from this study nevertheless provide evidence of the potential detrimental effect of PSS on standardized test performance. Given the unique longitudinal design of the NDHS, we had the opportunity to consider the baseline status of individuals who were eventually exposed to traumatic stressors, allowing for stronger causal inferences than those typically permitted within cross-sectional designs. Additional replication studies that include representative samples that are administered standardized college entrance tests as well as a clinical assessment of PTSD will allow findings to be applied to a broader population of college applicants. Future research will also benefit from consideration of the predictive validity of standardized academic assessments for those with PSS, including whether lower standardized test scores as a result of PTSD or PSS can accurately predict future academic performance.

Our findings have implications for the interpretation of standardized achievement assessment differences, particularly among students at high risk for PTSD and other psychiatric disorders that might affect test-taking ability. Differences in ability at the levels observed in this study do not inevitably imply biases sufficient to necessitate corrective action. However, given the prevalence of trauma exposure in the general population and the ubiquity of standardized assessments among college applicants, this study suggests that recognizing and understanding the potential additional disadvantages to which examinees with PSS are subject will be important to both examinees and educational counselors. In particular, prospective college students with PSS may benefit from counseling targeting coping strategies to help manage the negative emotional consequences of psychological trauma exposure and compensatory strategies to assist in test taking.

References

Abramson, D., & Garfield, R. (2006). *On the edge: Children and families displaced by Hurricanes Katrina and Rita face a looming medical and mental health crisis.* New York, NY: Columbia University Mailman School of Public Health.

Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22,* 47–76.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York, NY: Wiley.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Arbuthnot, K. N. (2005). *Stereotype threat as a source of differential item functioning (DIF): An examination of the standardized mathematics test-taking patterns and experiences of black adolescents* (Unpublished doctoral dissertation). University of Illinois at Urbana–Champaign.

Baron, J., & Norman, M. F. (1992). SATs, achievement tests, and high-school class rank as predictors of college performance. *Educational and Psychological Measurement, 52*(4), 1047–1055.

Blanchard, E., Jones-Alexander, J., Buckley, T., & Forneris, C. (1996).

Psychometric properties of the PTSD checklist (PCL). *Behaviour Research and Therapy, 34*(8), 669–673.

Brandes, D., Ben-Schachar, G., Gilboa, A., Bonne, O., Freedman, S., & Shalev, A. Y. (2002). PTSD symptoms and cognitive performance in recent trauma survivors. *Psychiatry Research, 110*(3), 231–238.

Brewin, C. R., Kleiner, J. S., Vasterling, J. J., & Field, A. P. (2007). Memory for emotionally neutral information in posttraumatic stress disorder: A meta-analytic investigation. *Journal of Abnormal Psychology, 116*(3), 448–463.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning test* (College Board Research Report No. 2000-1). New York, NY: College Entrance Examination Board.

Burton, N. W., & Wang, M. (2005). *Predicting long-term success in graduate school: A collaborative validity study* (Research Report No. RR-05-03). Princeton, NJ: Educational Testing Service.

Cheong, Y. F., & Raudenbush, S. W. (2000). Measurement and structural models for children's problem behaviors. *Psychological Methods, 5*(4), 477–495.

de Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach.* New York, NY: Springer.

Dohrenwend, B. P., Turner, J. B., Turse, N. A., Adams, B. G., Koenen, K. C., & Marshall, R. (2006, August 18). The psychological risks of Vietnam for U.S. veterans: A revisit with new data and methods. *Science, 313*(5789), 979–982.

Eccleston, J. (2007, March 20). *"Silent victims": What will become of Iraq's children?* Retrieved from http://edition.cnn.com/2007/WORLD/meast/03/19/iraq.children/index.html

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

FairTest. (2006). *SAT predictive validity fact sheet.* Retrieved from http://www.fairtest.org/facts/satvalidity.html

Gil, T., Calev, A., Greenberg, D., Kugelmass, S., & Lerer, B. (1990). Cognitive functioning in post-traumatic stress disorder. *Journal of Traumatic Stress, 3*(1), 29–45.

Gilbertson, M. W., Gurvits, T. V., Lasko, N. B., Orr, S. P., & Pitman, R. K. (2001). Multivariate assessment of explicit memory function in combat veterans with posttraumatic stress disorder. *Journal of Traumatic Stress, 14*(2), 413–432.

Gurvits, T. V., Gilbertson, M. W., Lasko, N. B., Tarhan, A. S., Simeon, D., Macklin, M. L., . . . & Pitman, R. K. (2000). Neurologic soft signs in chronic posttraumatic stress disorder. *Archives of General Psychiatry, 57*(2), 181–186.

Gurvits, T. V., Lasko, N. B., Schachter, S. C., Kuhne, A. A., Orr, S. P., & Pitman, R. K. (1993). Neurological status of Vietnam veterans with chronic posttraumatic stress disorder. *Journal of Neuropsychiatry and Clinical Neurosciences, 5*(2), 183–188.

Hart, J., Jr., Kimbrell, T., Fauver, P., Cherry, B. J., Pitcock, J., Booe, L. Q., . . . & Freeman, T. W. (2008). Cognitive dysfunctions associated with PTSD: Evidence from World War II prisoners of war. *Journal of Neuropsychiatry and Clinical Neuropsychiatry, 20*(3), 309–316.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47–77.

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine, 351*(1), 13–22.

Hoge, C. W., McGurk, D., Thomas, J. L., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in U.S. soldiers returning from Iraq. *New England Journal of Medicine, 358*(5), 453–463.

Holland, P. W., Hoffman, K., & Thompson, M. J. (2002). *An intergroup study of similarities and differences in the prediction of GRE general test scores* (Research Report No. RM-02–02). Princeton, NJ: Educational Testing Service.

Jones, W., Loe, S., Krach, S., Rager, R., & Jones, H. (2008). Automated Neuropsychological Assessment Metrics (ANAM) and Woodcock–Johnson III tests of cognitive ability: A concurrent validity study. *Clinical Neuropsychologist, 22*(2), 305–320.

Kabat, M. H., Kane, R. L., Jefferson, A. L., & DiPino, R. K. (2001). Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures. *Clinical Neuropsychologist, 15*(4), 498–507.

Kessler, R., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of *DSM-IV* disorders in the National Comorbidity Survey replication. *Archives of General Psychiatry, 62*(6), 593–602.

Kessler, R., Foster, C., Saunders, W., & Stang, P. (1995). Social consequences of psychiatric disorders I: Educational attainment. *American Journal of Psychiatry, 152*(7), 1026–1032.

King, D. W., King, L. A., & Vogt, D. S. (2003). *Manual for the Deployment Risk and Resilience Inventory (DRRI): A collection of measures for studying deployment-related experiences of military veterans.* Boston, MA: National Center for PTSD.

Krengel, M., White, R., Diamond, R., Letz, R., Cyrus, P., & Durso, R. (1996). A comparison of NES2 and traditional neuropsychological tests in a neurological patient sample. *Neurotoxicology and Teratology, 8*(4), 435–439.

Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1990). *Trauma and the Vietnam War generation: Report of findings from the National Vietnam Veterans Readjustment Study.* Philadelphia, PA: Brunner/Mazel.

Letz, R. (2000). *NES3 user's manual.* Atlanta, GA: Neurobehavioral Systems.

Macklin, M. L., Metzger, L. J., Litz, B. T., McNally, R. J., Lasko, N. B., Orr, S. P., & Pitman, R. K. (1998). Lower precombat intelligence is a risk factor for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 66*(2), 323–326.

Milliken, C. S., Auchterlonie, J. L., & Hoge, C. W. (2007). Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq War. *Journal of the American Medical Association, 298*(18), 2141–2148.

National Center for Education Statistics. (2005). *2000–01 through 2004–05 integrated postsecondary education data system "Institutional Characteristics Survey"* (IPEDS-IC:2000). Washington, DC: Author.

Pastor, D. A. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *Applied Measurement in Education, 16*(3), 223–243.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (Research Report No. RR-94-27). Princeton, NJ: Educational Testing Service.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (2nd ed.). Chicago, IL: University of Chicago Press.

Redlener, I. (2006, May 9). Orphans of the storm. *The New York Times.* Retrieved from http://www.nytimes.com

Reeves, D., Kane, R., Elsmore, T., Winter, K., & Bleiberg, J. (2002). *ANAM 2001 user's manual: Clinical and research modules* [Special Report No. NCRF-SR-2002]. San Diego, CA: National Cognitive Recovery Foundation.

Ruggiero, K. J., Del Ben, K., Scotti, J. R., & Rabalais, A. E. (2003). Psychometric properties of the PTSD Checklist–Civilian Version. *Journal of Traumatic Stress, 16*(5), 495–502.

Saigh, P. A., Mroueh, M., & Bremner, J. D. (1997). Scholastic impairments among traumatized adolescents. *Behaviour Research and Therapy, 35*(5), 429–436.

SAS Institute. (2003). *SAS OnlineDoc (Version 9)* [Software manual]. Cary, NC: Author.

Schell, T. L., & Marshall, G. N. (2008). Survey of individuals previously deployed for OEF/OIF. In T. Tanielian & L. H. Jaycox (Eds.), *Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery* (pp. 87–115). Santa Monica, CA: RAND.

Schmitt, A. P., & Dorans, N. J. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement, 27*(1), 67–81.

Schwab-Stone, M. E., Ayers, T. S., Kasprow, W., Voyce, C., Barone, C., Shriver, T., & Weissberg, R. P. (1995). No safe haven: A study of violence exposure in an urban community. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*(10), 1343–1352.

Schwartz, D., & Gorman, A. H. (2003). Community violence exposure and children's academic functioning. *Journal of Educational Psychology, 95*(1), 163–173.

Sheu, C.-F., Chen, C.-T., Su, Y.-H., & Wang, W.-C. (2005). Using SAS PROC NLMIXED to fit item response theory models. *Behavior Research Methods, 37*(2), 202–218.

Smith, T. C., Ryan, M. A., Wingard, D. L., Slymen, D. J., Sallis, J. F., & Kritz-Silverstein, D. (2008). New onset and persistent symptoms of post-traumatic stress disorder self reported after deployment and combat exposures: Prospective population based US military cohort study. *British Medical Journal, 336*, 366–371.

Tanielian, T., & Jaycox, L. H. (Eds.). (2008). *Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery.* Santa Monica, CA: RAND.

Thienkrua, W., Cardozo, B., Chakkraband, M., Guadamuz, T., Pengjuntr, W., Tantipiwatanaskul, P., et al. (2006). Symptoms of posttraumatic stress disorder and depression among children in tsunami-affected areas in southern Thailand. *Journal of the American Medical Association, 296*(5), 549–559.

Vasterling, J. J., & Brailey, K. (2005). Neuropsychological findings in adults with PTSD. In J. J. Vasterling & C. R. Brewin (Eds.) *Neuropsychology of PTSD: Biological, Cognitive, and Clinical Perspectives* (pp. 178–207). New York, NY: Guilford.

Vasterling, J. J., Brailey, K., Constans, J. I., Borges, A., & Sutker, P. B. (1997). Assessment of intellectual resources in Gulf War veterans: Relationship to PTSD. *Assessment, 4*(1), 51–59.

Vasterling, J. J., Duke, L. M., Brailey, K., Constans, J. I., Allain, A. N. J., & Sutker, P. B. (2002). Attention, learning, and memory performances and intellectual resources in Vietnam veterans: PTSD and no disorder comparisons. *Neuropsychology, 16*(1), 5–14.

Vasterling, J. J., Proctor, S. P., Amoroso, P., Kane, R., Gacksletter, G., Ryan, M. A. K., & Friedman, M. J. (2006). The Neurocognition Deployment Health Study: A prospective cohort study of Army soldiers. *Military Medicine, 171*(3), 253–260.

Weathers, F. W., Huska, J. A., & Keane, T. M. (1991). *The PTSD Checklist, Civilian version (PCL-C).* Boston, MA: National Center for PTSD.

Winter, G. (2005, January 16). From combat to campus on the GI Bill. *The New York Times.* Retrieved from http://www.nytimes.com

Zwinderman, A. H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika, 56*(4), 589–600.

Zwinderman, A. H. (1997). Response models with manifest predictors. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 245–256). New York, NY: Springer.

Received April 22, 2008
Revision received July 1, 2009
Accepted July 16, 2009 ■